DOCUMENT RESUME

ED 464 924 TM 034 156

AUTHOR Frauenholtz, Todd
TITLE College Instructors' Evaluation of Students: Thoughts on the Student Evaluation Standards.
PUB DATE 2002-04-00
NOTE 9p.
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Research (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; *College Students; *Evaluation Methods; Higher Education; Standards; *Student Evaluation

ABSTRACT

This paper presents the results of four students' observations of instructors' evaluation practices. The four students (two undergraduates and two graduate students) rated an instructor's evaluation of students using a draft of the Student Evaluation Standards (SES) of the Joint Committee at Western Michigan University. Each student wrote a paper that became part of a family of studies presented at the 10th Annual Consortium for Research on Educational Accountability and Teacher Evaluation National Institute. This paper summarizes those papers, with a brief quantitative overview of the students' observations. Two of the students applied the SES to courses in which they thought the instructor was exemplary, while the other two applied the SES to courses in which they thought evaluation was frustrating and in need of improvement. Students were given a draft copy of the SES, and the ratings generated by the four students were entered into a spreadsheet according to the four attributes of sound evaluation practice (Propriety, Utility, Feasibility, and Accuracy). The ratings of instructors' practices provided by students participating in this study show that there is room for improvement in the field of student evaluation. The SES are capable of providing a foundation on which educators can build solid student evaluation policies and practices. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 924

College Instructors' Evaluation of Students: Thoughts on the Student Evaluation Standards

Todd Frauenholtz
Bemidji State University
Bemidji, MN

This paper presents the results of four students' observations of instructors' evaluation practices. The students were: an undergraduate pursuing a degree in Spanish, an undergraduate pursuing a degree in engineering mechanics, a graduate student pursuing a master's degree in divinity, and a graduate student pursuing a Ph.D. in mathematics education. These four students were asked to choose one instructor from the 2000-2001 academic year and rate that instructors' evaluation of students using a draft of the Student Evaluation Standards (SES) as categories to be judged. Each student chose a course in his or her major content area.

Each student wrote a paper that was part of a family of studies presented at the tenth annual Consortium for Research on Educational Accountability and Teacher Evaluation National Evaluation Institute in July of 2001. This paper presents a summary of those four papers and a brief quantitative overview of the four students' observations.

Of the four students, two chose to apply the SES to a course in which they believed the instructor was an exemplary model of what a college instructor should be; perhaps their "favorite" instructor. The other two students chose courses where they found the evaluation procedures to be frustrating and in need of improvement. This two by two design occurred merely by chance, but it does illustrate the variation that occurs in classrooms. This paper is, first, an attempt to determine how well college instructors' practices met the standards and second an examination of a draft version of the SES to ascertain the extent to which these standards could be useful in evaluating the practices an instructor uses to evaluate students.

Procedures

Students were provided with a draft copy, dated December 2000, of the SES and asked to apply the standards to the practices of one instructor upon completion of the course. Each student went through the standards and commented on those that stood out as having been well or poorly addressed by the instructor of the chosen course. For this paper each student was asked to rate the instructor's practices on each of the 29 standards using a ten-point scale. A rating of one meaning the instructor poorly addressed the area and a rating of ten indicating that the instructor's practices in this area were exemplary of the type of conduct the standards suggest. If the standard was not observed it was denoted by the letters NO.

The ratings generated by the four students were entered into a spreadsheet. The columns represented the students and the rows were the individual standards grouped according to the four attributes of sound evaluation practice (Propriety, Utility, Feasibility, and Accuracy). Means were calculated for each student and each standard. The results are presented in the following pages.

TM034156

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
T. Frauenholtz
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

P Standards

The Propriety standards, and the associated ratings by each student, are presented in Table 1. As a group, the Propriety standards had the highest average ranking, 6.30 on the ten-point scale described earlier. The lowest rated standard in the Propriety group was the Conflict of interest (P7) standard. This was an area that was either not observed or was done poorly with the exception being the divinity student's experiences. This standard, P7, received the fourth lowest ranking of all 29 standards. Instructors were not rated low on any other standards in this group when the students' ratings were averaged.

Table 1
Students' ratings of Propriety Standards

Standard	Mean	Math. Ed.	Divinity	Spanish	Engineering
P1 – Service orientation	7.00	10	5	5	8
P2 – Appropriate policies/procedures	6.33	8	3	NO	8
P3 – Access to evaluation info.	7.67	5	9	NO	9
P4 – Treatment of students	6.00	10	7	2	5
P5 – Rights of students	7.75	9	8	5	9
P6 – Balanced evaluation	6.00	8	3	5	8
P7 – Conflict of interest	3.33	2	6	2	NO

Access to evaluation information (P3) was the fourth highest rated standard by the students considering all standards and P5 (Rights of students) was the third highest. In P3 the engineering and divinity students had the most positive experiences and in P5 the mathematics education and engineering students provided the highest marks with the divinity student also giving a high mark in this area. The mathematics education and engineering students both found their instructors to have done very well in the Propriety area overall and the Spanish student found these standards not well addressed in her class.

U Standards

The Utility standards were given the lowest ranking as a group, 4.71 on the ten-point scale. The rankings are presented in Table 2 and generally show consistently low marks. The engineering student found exception to this giving her instructor consistently high marks except in the Follow-up standard (U7). This was the lowest standard in this area and the second lowest of all the SES rated by the four students.

Table 2
Students' ratings of Utility Standards

Standard	Mean	Math. Ed.	Divinity	Spanish	Engineering
U1 – Constructive orientation	4.00	4	2	2	8
U2 – Defined uses and users	6.00	3	7	5	9
U3 – Information scope	5.50	7	4	3	8
U4 – Evaluator qualifications	5.50	NO	NO	4	7
U5 – Explicit values	4.00	4	4	NO	NO
U6 – Effective reporting	5.25	3	4	5	9
U7 – Follow-up	2.75	2	6	1	2

In the Utility group the highest rated standard was U2 (Defined uses and users) with an average rating of six on the one to ten scale. This was not enough to rank it among the top scoring standards overall. The engineering student again found the best practices by her instructor and the Spanish student reported low marks.

F Standards

The Feasibility standards were the second highest, 6.08 on the ten-point scale, rated group (behind the Propriety standards) and results are provided in Table 3. In this area the engineering student delivered the lowest marks with the Spanish student also providing low ratings while the mathematics education and divinity students observed better coverage of the Feasibility standards. All three Feasibility standards were clumped closely together but F2 (Political viability) had slightly lower marks.

Table 3
Students' ratings of Feasibility Standards

Standard	Mean	Math. Ed.	Divinity	Spanish	Engineering
F1 – Practical orientation	6.25	8	9	5	3
F2 – Political viability	5.75	10	7	3	3
F3 – Evaluation support	6.25	7	7	5	6

The other two Feasibility standards (F1 – Practical orientation and F3 – Evaluation support) tied for top honors in the group. None of the Feasibility standards were individually noteworthy but the mathematics education and divinity students rated their instructors well while the Spanish and engineering students provided low marks.

A Standards

Table 4 presents the rankings reported for the Accuracy standards as observed by the four students. It is noteworthy to this author that the Accuracy group had the most standards as well as the highest proportion of "Not Observed" responses. Additionally, this group earned the second lowest group mean, 5.81 on the ten-point scale, of the four areas but also provided the largest range of mean scores. The third lowest score for a standard overall was given to A5 (Documented procedures) and the overall lowest score goes to A12 (Metaevaluation) with a score of 2.0; this standards was also not observed by two students.

Table 4
Students' ratings of Accuracy Standards

Standard	Mean	Math. Ed.	Divinity	Spanish	Engineering
A1 – Validity orientation	6.00	8	2	4	10
A2 – Justified conclusions	5.00	7	3	3	7
A3 – Defined expectations (students)	6.75	9	5	3	10
A4 – Context analysis	6.00	7	NO	3	8
A5 – Documented procedures	3.00	4	2	3	NO
A6 – Defensible information sources	6.25	10	6	3	6
A7 – Reliable information	6.00	8	2	NO	8
A8 – Handling information & q.c.	8.00	NO	8	NO	NO
A9 – Analysis of quantitative info	9.00	8	NO	NO	10
A10 – Analysis of qualitative info	6.00	NO	NO	NO	6
A11 – Bias identification and mgmt.	5.67	NO	5	3	9
A12 – Metaevaluation	2.00	NO	2	2	NO

The Accuracy group also provided the two highest rated standards. A9 (Analysis of quantitative information) had the highest rating overall, though it went unobserved by two of the students. Standard A8 (Handling information and quality control) had the second highest rating overall, but only one student observed it.

Highest scoring standards

There were several standards that students rated as having been done well by their instructors, these are presented in Table 5. The two highest rated standards went unobserved by several of the students. This may be interpreted by some as an indication that things were not done well; however upon examination of the two standards (A9 – Analysis of quantitative information and A8 – Handling information and quality control) it may appear reasonable that these are things which students should not observe due to confidentiality issues. On the other hand, perhaps students have become so accustomed to not knowing how they are being evaluated that they don't expect to be informed. Taking this into consideration, it is reassuring to know that what students are able to observe of instructors' handling and analysis of information students perceive as being done well. It would be more reassuring to know that instructors are evaluating students well. Instructors should be able to explain **how** quantitative data were analyzed without violating confidentiality.

Table 5
Students' highest rated standards

Standard	Mean	Math. Ed.	Divinity	Spanish	Engineering
A9 – Analysis of quantitative info	9.00	8	NO	NO	10
A8 – Handling information & q.c.	8.00	NO	8	NO	NO
P5 – Rights of students	7.75	9	8	5	9
P3 – Access to evaluation info.	7.67	5	9	NO	9

The final two standards on which instructors were rated highly by the students are P5 (Rights of students) and P3 (Access to evaluation information). Rights of students was rated well by all students with the exception of the Spanish student while access to evaluation was rated well by the divinity and engineering students and unobserved by the Spanish student.

It seems nearly impossible to set one or two standards as being of a higher priority due to the wide variation of circumstances. What is vital in one setting may be significantly less important in another. The draft document of the SES contains a Functional Table of Contents that helps users determine which standards might be most relevant to several general situations, but with the uniqueness of every educational setting it seems unreasonable for this author to determine where these four standards should rank on the continuum of most or least important areas of student evaluation. In an ideal setting the most important standards for a given situation are the ones that instructors would do well.

Lowest scoring standards

While some standards were rated higher than most, there are also those that were rated lower. The low rated standards are presented in Table 6. The lowest rated standard was A12 (Metaevaluation) which was not observed by two students and rated low by the other two. Three of the four students found U7 (Follow-up) to have been done poorly by their instructor.

Table 6
Students' lowest rated standards

Standard	Mean	Math. Ed.	Divinity	Spanish	Engineering
A12 – Metaevaluation	2.00	NO	2	2	NO
U7 – Follow-up	2.75	2	6	1	2
A5 – Documented procedures	3.00	4	2	3	NO
P7 – Conflict of interest	3.33	2	6	2	NO

Three of the students rated the A5 (Documented procedures) standard low and one student did not observe anything that addressed this standard. In this particular standard, it seems unlikely that confidentiality plays a role here as a rationale for not documenting evaluation procedures. Conflict of interest (P7) is the final standard that was rated low by the students. Here two students reported the conflict of interest standard as going unmet while one student seemed to rate it about average and the fourth student did not observe any conflict of interest issues.

It is interesting to compare the "NO" ratings in Table 5 with those of Table 6. In the first table the Spanish student contributed most of the "NO" ratings. This student rated her instructor's

evaluation practices the lowest of the four students. In the latter table, the engineering student provided the most "NO" ratings. The engineering student rated her instructor's evaluation practices the highest out of the four participating students. Henceforth, it seems that high ratings come from two things: first, high ratings from some students providing high marks and second, a lack of low ratings from other students. Doing student evaluation well involves not only doing the right things but avoiding those things that are detrimental.

Overall evaluation of the standards document

The four students generally agreed that the SES document was clearly written and easy to use. There were several instances where a specific standard was not observable by the student. It is possible that the instructor was doing some activity that met the letter of the standards; however, students were not privy to every move an instructor made. Students found that the standards were general enough that instructors could meet the standards by different means. For instance, P2 was rated an 8 by both the mathematics education and engineering students. In the case of the engineering student, she felt that the syllabus very clearly outlined the policies and procedures that would be used in the course. The mathematics education course read an article the first day of class. This article brought out issues of grading which lead to a class discussion where the instructor informed the class what he would be doing. Both instructors addressed the standard, but in quite different ways.

Summary and conclusions

The ratings of instructors' evaluation practices provided by students participating in this study indicate that there is room for improvement in the field of student evaluation. The SES are capable of providing a foundation upon which educators can build solid student evaluation policies and practices.

How well college instructors' evaluation practices were rated depends on which students are considered. The mathematics education and engineering students provided the highest ratings for instructor evaluation practices, but even their numbers indicate that there is room for improvement. The Spanish student's experience seems to suggest that evaluation practices, in at least one classroom, are in need of certain improvement. There appears to be room for improvement in all classrooms when looking at the averages from the four students participating.

All the standards and student ratings of instructors are presented in Table 7.

Table 7
Ratings from all students and all standards

Standard	Average	Math. Ed.	Divinity	Spanish	Engineering
P1 - Service orientation	7.00	10	5	5	8
P2 – Appropriate policies and procedures	6.33	8	3	NO	8
P3 - Access to evaluation information	7.67	5	9	NO	9
P4 - Treatment of students	6.00	10	7	2	5
P5 - Rights of students	7.75	9	8	5	9
P6 - Balanced evaluation	6.00	8	3	5	8
P7 - Conflict of interest	3.33	2	6	2	NO
P Average:	6.30				
U1 - Constructive orientation	4.00	4	2	2	8
U2 - Defined uses and users	6.00	3	7	5	9
U3 – Information scope	5.50	7	4	3	8
U4 - Evaluator qualifications	5.50	NO	NO	4	7
U5 - Explicit values	4.00	4	4	NO	NO
U6 - Effective reporting	5.25	3	4	5	9
U7 - Follow-up	2.75	2	6	1	2
U Average:	4.71				
F1 - Practical orientation	6.25	8	9	5	3
F2 - Political viability	5.75	10	7	3	3
F3 - Evaluation support	6.25	7	7	5	6
F Average:	6.08				
A1 - Validity orientation	6.00	8	2	4	10
A2 - Justified conclusions	5.00	7	3	3	7
A3 - Defined expectations for students	6.75	9	5	3	10
A4 - Context analysis	6.00	7	NO	3	8
A5 - Documented procedures	3.00	4	2	3	NO
A6 - Defensible information sources	6.25	10	6	3	6
A7 - Reliable information	6.00	8	2	NO	8
A8 - Handling information and quality control	8.00	NO	8	NO	NO
A9 - Analysis of quantitative information	9.00	8	NO	NO	10
A10 - Analysis of qualitative information	6.00	NO	NO	NO	6
A11 - Bias identification and management	5.67	NO	5	3	9
A12 – Metaevaluation	2.00	NO	2	2	NO
A Average:	5.81				
Overall averages	5.63405	6.708	5.04	3.4545	7.33

It could also be argued that unobserved standards are indicators of evaluation not being done well. One of the themes I found in the SES is that students should be informed of what is expected of them and how their performance will be evaluated; in short, no surprises from the instructor. Looking at the number of standards that students reported as not observed, there is some consistency regardless of the overall averages from the different students. The students, from left to right, had 5, 4, 7, and 5 incidents of NO's in Table 7. This indicates that there are some things that remain shrouded from the students both in classrooms where student evaluation

was perceived as well done and in classrooms where the evaluation appeared poorly done. Additionally, with the many unobserved standards, I am lead to believe that, while the mean scores are accurate, they must be interpreted with some caution. Several of the standards received NO's from more than one student. These are certainly standards that people seriously considering student evaluation should consider carefully and ask why it is that students report not observing these standards. Is the issue confidentiality, poor evaluation practice, or something else?

Additionally on the NO topic, "Not Observed" was not factored into the numerical value when computing averages, if a rating of one, the lowest possible value, were substituted some standards' scores would change dramatically. It seems that if the instructors do not appear to be doing something to address the standards, then scoring them low on that particular standard would be fair. On the other hand, there are issues to which students should not have access and instructors' efforts at protecting confidentiality, or other delicate matters should be recognized with an appropriately high score. Since it is not possible to determine which category best fits the "NO" ratings they remain excluded from calculation of the mean scores in this study.

When considering the usefulness of the standards document, the students provided generally positive comments. The SES were well organized and easy to read. The organization was well done by using the four important attributes of sound evaluation practice: propriety, utility, feasibility, and accuracy. Each standard has a title, brief description, definition of the standard, a conceptual overview, guidelines, common errors, illustrative cases, and supporting documentation. Students generally found the SES easy to understand; however, there were many standards rated as "Not Observed" by students. Some of the students claim these are things that they are accustomed to not seeing – no instructors do them. For example, exactly how an instructor grades assignments/exams or how scores are recorded and analyzed into a final letter grade is something students rarely learn. The students here appear comfortable accepting, on good faith, that these are areas where the instructor "knows what s/he is doing." The SES exist, in part, because too many instructors did NOT know what they were doing! A student also raised the point that students using the SES alone will not improve student evaluation, there needs to be some involvement of the instructor in this appraisal of evaluation practices for improvement to occur. Students may have a role in this appraisal process but they cannot be the entire process on their own. In conclusion, evaluation plays a large role in classroom instruction and doing evaluation well is a worthy goal that the SES attempt to advance.